The Milbank Quarterly
MS#18-126
Received 9/11/2018

Ranking hospitals based on preventable hospital death rates: A Systematic Review with implications for both direct measurement and indirect measurement through standardized mortality rates.

ABSTRACT

Objectives: There is interest in monitoring avoidable or preventable deaths measured directly or indirectly through standardized mortality rates (SMRs). We reviewed studies that use implicit case note reviews to estimate the range of preventable death rates observed, the measurement characteristics of those estimates, and the measurement procedures used to generate them. We comment on the implications for monitoring SMRs and illustrate a way to calculate the number of reviews needed to establish a reliable estimate of preventability of one death or the hospital preventable death rate.

Design: Systematic review of the literature supplemented by re-analysis of authors previously published and un-published data and measurement design calculations.

Data source: Searches in PubMed, MEDLINE (OvidSP) and Web of Knowledge in June 2012, updated December 2017.

Eligibility criteria: Studies of hospital-wide admissions from general and acute medical wards where preventable deaths rates are provided or can be estimate and which can provide inter- observer variations.

Results: Twenty-four studies were included from 1983-2017. Recent larger studies suggest consistently low rates of preventable deaths (3.0-6.5% since 2012). Reliability of a single review for distinguishing between individual cases with regard to the preventability of death had a Kappa rate of 0.27-0.50 for deaths and 0.24-0.76 for adverse events. A Kappa of 0.35 would require an average of 8-17 reviews of a single case to be precise enough to have confidence about high stakes decisions to change care procedures or impose sanctions within a hospital as a result. No study estimated the variation in preventable deaths across hospitals, although we were able to re-analyse one study to obtain an estimate. Based on this estimate, 200-300 total case-note reviews per hospital could be required to reliably distinguish between hospitals.

The studies display considerable heterogeneity: 13/24 studies defined preventable with a threshold of ≥4 in a six-category Likert scale; 11/24 involved a two-stage screening process with nurses at the first stage and physicians at the second. Fifteen studies provided expert clinical review support for reviewer disagreements, advice, or quality control. A 'generalist/internist' was the modal physician specialty for reviewers and they received 1-3 days of generic tools orientation and case-note review practice. Methods did not consider the influence of human or environmental factors.

Conclusions: The literature provides limited information about the measurement characteristics of preventable deaths that suggests substantial numbers of reviews may be needed to create reliable estimates of preventable deaths at the individual or hospital level. Any operational program would require population specific estimates of reliability. Preventable death rates are low, which is likely to make it difficult to use SMRs based on all deaths to validly profile hospitals. The literature provides little information to guide improvements in the measurement procedures.

Systematic review registration: The systematic review was conceived prior to PROSPERO, and so has not been registered.

KEY WORDS: Avoidable, Preventable, hospital deaths, hospital mortality, systematic review, variation

SUMMARY BOX

What is Already Known on This Topic:

- Numerous studies of adverse events including preventable deaths using implicit case-note reviews by clinicians have taken place in recent years, but no systemic reviews have aimed to summarise the estimates or the variations in methodologies used to derive these estimates.
- Several health systems have suggested measuring preventable death rates of hospitals for comparison across time or in league tables and the use of standardized mortality rates to profile hospitals presumes differences in preventable deaths. The methodology and optimal review numbers per case note or hospital have not been explored.

What This Study Adds:

- Estimates for preventable rates using implicit case-note reviews by clinicians are generally low with poor to fair agreement between reviewers seen across studies.
- Low preventable death rates suggest that those using SMRs to rank hospitals will need extraordinarily good risk adjustment models to be useful.
- Based on limited information in the studies substantial numbers of reviews would be required for learning from individual cases or profiling hospitals.
- Population and hospital system specific data on variation across hospital/providers to be compared for preventable deaths or adverse events is needed to establish adequate numbers of reviews in order to design a measurement procedure that can distinguish between the units.

INTRODUCTION

Standardized mortality ratios(SMRs) for hospitals are currently used as an indicator of institutional quality and compared between hospitals to identify "outliers".(1) The rationale for their use is that they are a proxy for "excess" or preventable deaths, but there are compelling arguments that any signal (preventable death) will be obscured by the noise (all other unavoidable deaths).(2, 3) Some policy-makers are considering using direct measurements of preventable mortality, rather than trying to infer it indirectly from SMRs, as with the Summary Hospital-level Mortality Indicator (SHMI) used in the English NHS.(4-7) For example, the NHS in England has instituted a system of mandatory physician review of deaths in hospital in order to establish (and publish) the number of preventable deaths for local Trust use and learning from mistakes.(8, 9) A direct measurement of preventable death is also an obvious way to validate the widespread use of SMRs as a way to measure the quality of care delivered to people prior to their death.

However, preventable death, as well as preventable adverse events more broadly, can only be directly measured by the judgement of expert clinical observers who retrospectively review case-notes. Such judgement-based assessments generally have low reliability, meaning that they lack consistency across repeated reviews. Thus, current or future policy and research agendas that propose measuring any preventable adverse events, and specifically preventable mortality, should push us to define, and if possible improve the measurement characteristics of those estimates. Only then can we use case note review measurements in research to validate SMRs or consider trying to design operational systems

for learning from adverse events within hospitals or even comparing preventable deaths between hospitals, possibly augmenting or even replacing comparisons by means of SMRs.

To this end, we conducted a systematic review firstly to summarise data from existing studies about the number of avoidable deaths and the measurement characteristics of those estimates and apply these in order to determine the number of reviews that would be needed to establish a reliable preventable death estimate at the individual or hospital level. Secondly, to summarise the heterogeneity between the measurement procedures used in these studies including reviewer characteristics, selection and training factors to assess whether there are potential opportunities to improve the reliability of the measurement procedure. This is the first review of methods to measure preventable mortality rates.

METHODS

Literature search

An initial search was undertaken within the PubMed and ISI Web of Knowledge in 2010. This was updated and supplemented by a broader search in MEDLINE (OvidSP) incorporating a wider range of terms covering preventability and errors, deaths and AEs, hospitals and case-note reviews in June 2012 and December 2017 (see Appendix 1). After our last search and before finalising this manuscript, we were made aware of two studies that met our inclusion criteria.(10-12) These studies are included in our review to ensure our findings remain up-to-date. Reference lists of included studies were also hand-searched to find additional articles.

Study selection

The inclusion criteria were studies which (a) evaluate the preventability of hospital deaths (deaths primarily from general and acute medical wards) or preventable AEs contributing to death from a hospital-wide sample or primarily from general and acute medical wards; (b) provide a quantitative estimate of preventability of death or allow this to be calculated; (c) incorporate an implicit review process that elicits the reviewer's own expert judgment, in reaching the conclusion about preventability. Only articles published in English were considered. Two reviewers (authors KKG, PJC or APT) independently examined titles and abstracts retrieved from literature searches and selected studies for inclusion.

Disagreements were resolved by consensus after retrieval of full-text and further discussions with a third reviewer (Y-FC). The review protocol was not submitted to PROSPERO as the review process was initiated before the establishment of PROSPERO.

Data extraction and synthesis of evidence

Two reviewers (KKG, Y-FC, PJC, or APT) extracted data from the selected studies, including all data tabulated in Tables 1-3. The characteristics and findings of included studies were tabulated and summarized in a narrative form. We did not plan to pool results across studies given the underlying differences in settings and methods between the studies. Where data was missing, authors were written to and details obtained.

Number of reviewers required for a reliable measurement

Reliability increases with a measurement procedure that makes multiple independent measurements by different reviewers and averages them. Most reports of the reliability of case-note review give a number that describes the ability of a single review of any one case note to distinguish between patients who died with respect to whether the death was preventable. In appendix 2, we describe one method to calculate how the reliability increases with the number of reviews.

These commonly reported reliability estimates, which describe the ability to distinguish between case-notes of patients who died, can quantify the confidence with which one can act on the presumption that a specific avoidable death had occurred, such as by investing in doing a root cause analysis to establish proximate causes, or possibly for establishing legal liability or compensation *for an individual case*. However, such reliability estimates tell you nothing about determining the performance of *different providers*, such as different hospitals. In any measurement a key determinant of reliability is the variation across the things you want to distinguish between and thus to distinguish between hospitals requires an estimate of the variation of preventable death rates *across* hospitals.

No study was found to have published an estimate of this quantity despite its critical relevance to any policy making with respect to preventable deaths. We were able to re-analyse data from one study with 22 hospitals to produce the variance estimates required to

make a provisional “best available” calculation of the optimal number of reviews per case and per hospital required to produce a reliable estimate of the hospital preventable death rate (Appendix 2).(13) Only one other study had quantified hospital variation, finding a hospital variance estimate similar in magnitude to the one we estimated, however this was for a more global measurement of preventable adverse events that include deaths.(14)

RESULTS

Article retrieval and inclusion (Figure 1)

The electronic searches yielded 663 records after duplicates had been removed. Citation search of included studies identified six additional articles. In all, 37 articles (representing 24 studies) were included.(10-46) The characteristics of included studies are shown in Table 1. The study selection process and reasons for exclusion are summarised in Appendix 3. In 23 studies, data from publications were inadequate. From these, 14 of the authors written to responded.

[TABLE 1]

Thirteen studies (10-13, 21, 26-33, 36, 38) focused exclusively on the assessment of preventable death. Eleven studies (14-20, 22-25, 34, 35, 37, 39-46) primarily aimed to evaluate preventable AEs, which included, but were not confined to deaths. All except for two studies were in high income-countries conducted between 1984 and 2015 They involved a median of 13 hospitals (IQR = 24.25) and 1,068 case-notes (range 10 to 30,121).

Methods for assessing preventable deaths and preventable adverse events contributing to deaths

The majority of the published studies did not have enough details to obtain the information required for this review and unpublished data was obtained through author communications. Through writing to the authors, we obtained additional data from 13 of the 24 studies. These are summarised below in Table 2 and Appendices 4 and 5.

[TABLE 2]

a) *Tools and stages of review (Table 2 and Appendices 4 and 5)*

A plurality of the studies (10/24) followed the method of the Harvard Medical Practice Study,(22) which in turn was based on an approach called structured implicit physician review developed by RAND in the 1980s.(47) This measurement procedure includes an initial screening of patient notes to identify cases in which it is more likely that an adverse event might have occurred. The other studies provided a varied amount information on methodology and therefore we wrote to authors for details. These are summarised in Table 2 and Appendices 4 and 5.

In structured implicit case-note review the structured component guides the reviewer systematically and more or less temporally through the hospital admission, asking them to focus and rate specific elements of the patient's care in sequence before making an overall judgment about the quality of care.(48) The 'implicit' component is inherent in the summary judgements produced by the reviewer about the case, as opposed to generating a score based on a checklist, as well as the exercise of professional situational judgement in deciding whether deviations from ideal processes represent an error or are appropriate in the clinical context. A 'non-structured implicit' review has been found to be less reliable in estimating hospital quality of care, presumably due to the less standardized approach for navigating a record and building up to an overall rating.(49, 50)

In our sample, most studies used a kind of structured implicit (or criterion-based implicit) review pro forma. Although the details of the structured component varied, in all cases adopting structured implicit review, the 'structured' component required the reviewer to review and make quality judgements over phases of care (such as diagnostic or treatment phase). The reviewer was often asked to write explicit comments about areas of concern (as free-flow text) for each phase, and finally to score quality for each phase of care.

The decision on preventability was made on a scale applying implicit judgement of the physician reviewers. The majority (15/24) of the studies used a 6-category grading system (Likert scale) to classify the preventability of deaths and/or AEs.(10, 14-20, 22-25, 32-35, 37-46) The categories were often collapsed into a binary outcome. Deaths (and/or AEs) with a more than 50/50 preventable on balance of probability were considered preventable in most studies. Three studies (11-13, 31) used a continuous scale 0-100 probability of

preventability comparing it to the Likert scale; the 0-100 scale was found to have the same constructs and imparts comparable information to the Likert scale.(13)

Only five studies noted an attempt to anonymise the patient and hospital identifiers in case-notes (13, 23-25, 31-33) to prevent bias during reviews. No study blinded the reviewers to the outcome in these samples selected on the basis of death as the outcome.

b) Reviewer Selection and Training (Table 2 and Appendices 4 and 5)

All reviewers were external to the institutions from which case-notes were derived to reduce internal institutional bias. For reviewer selection, seven studies did not have a first stage screening process and deployed only physicians for these reviews.(13, 14, 18, 19, 27-29, 32, 33, 36, 44-46) Fifteen studies used two-stages, a screening process that involved mainly nurses at the first stage and exclusively physician reviewers at the second stage.(10, 14-26, 31, 34, 35, 38-46) Seven studies used an expert reviewer physician: in six studies they settled disagreements between the physician reviewers (14, 16-19, 22-25, 37, 43-46); while in one they were for quality control purposes.(39-42)

The required reviewer experience (where recorded) varied widely across the studies in both nurses and physicians. For physicians, regular handling of case-notes, a lengthy period of clinical work (i.e. often >5 years clinical/reviewing experience), postgraduate education and independent accreditation were used. For example, in the US studies, these were board-certified with the general preference for generalists/internists.(10, 21, 22, 43) The UK studies used specialties across general medicine and intensive care consultants.(13, 32, 33, 38) Eight studies deployed general physicians,(11-19, 22-25, 37, 43-46) and in seven of these a panel of specialists were available to advise individual reviewers when required.(11, 12, 14-19, 22-25, 37, 43-46)

Various forms of reviewer training and support were provided. The training duration varied between 1-3 days. Nurses and physicians had the same training in eight studies.(14-20, 22-25, 35, 37, 44-46) Eleven studies (11/24) are explicit about the exposure to case-notes during the training.(10, 13, 14, 18-28, 36, 37, 44-46) Six studies did not disclose reviewer

training information. Where enough details were available, training did not define preventability, but rather offered clinicians an opportunity to understand the aims, merits and some caveats (e.g. hind-sight bias (51, 52)) of the case-note review process, familiarise them with the pro forma for data extraction and to exchange views on approaches to difficult cases after practicing the review on one or more case-notes.(13, 16, 17, 19, 20, 22-24, 26, 31-34, 46)

Estimated preventable mortality (Table 3)

The proportion of deaths judged to be preventable depends upon the cut-off threshold used in the Likert scale. Preventable mortality rates as a proportion of 'all admissions' were estimated between 0.07% and 4.62% (Table 3). Most reports were below 0.7%; the 2.27% reported in Brazil (34) and the two Dubois studies (26-28) were exceptionally high. Preventability rates as a proportion of 'all deaths' were estimated between 0.47% and 29%. (10-13, 16-19, 21, 26, 28-34, 36, 38, 43-46, 53, 54) The estimates become more consistent when considering only more recent years (2008 to 2017) and high-income countries (3.0–6.0%). One study chose to estimate preventability at the lowest threshold, namely 'any probability that the death could have been prevented (e.g. ≥2 out of 6).'(23, 24)
The studies that evaluated preventability of all AEs as a proportion of 'all admissions' reported generally higher but widely variable figures for AEs, ranging from 1.02% (22) to 11.65%(16, 17), and preventable AEs as a proportion of 'all AEs' ranging from 3.96% (22) to 70.1%.(37)

Inter-rater reliability (Kappa Statistic) (Table 1)

Agreements or inter-rater agreements in assessing preventability of death or AEs contributing to death were reported in 20 studies.(10-15, 18, 19, 21-28, 30-33, 35-46) Cohen's Kappa statistics are reported for 18 of these 20 studies.(10-13, 15, 21-26, 30-33, 35-43) Cohen's Kappa is a statistic that was developed to measure the agreement between subjects taking into consideration the agreement that occurs by chance,(55) although for these ordinal measures the intra-class correlation would probably be preferred.(56, 57) Reported preventable mortality agreements were moderate and ranged from a Kappa of 0.27 (13) to 0.50.(31) Six studies reported reliabilities for assessing the preventability of all AEs (including deaths) (14, 18, 19, 22, 26, 35, 37, 43-46) and the reported AE Kappa ranged

from 0.19 to 0.76.(15, 20, 22, 35, 37, 39-43) Six studies report a Kappa for the preventability of AEs as judged by physicians, between 0.33 and 0.83.(11, 12, 20, 35, 37, 39-43) One study reports a Kappa for nurses of 0.40.(14, 18, 19, 44-46)

No data was found on effects of review selection, characteristics or training on the judgement of preventability by the reviewers.

Calculating the optimal number of reviews and reviewers per case note to estimate preventable death per case-note and per hospital

The range of reliability reported for the ability of a single review to distinguish between cases was 0.27 to 0.50. At a representative level of reliability of 0.35 for a single review, an average of 8 reviews per case-note would be required to achieve a reliability of 0.8 when distinguishing between cases. Seventeen reviews per case would be required to achieve a reliability of 0.9, a level often recommended for testing with high stakes consequences. However, any given operational program would have to determine the reliability of their measurement procedure in their population to figure out the number of cases needed to review.

About 200-300 total reviews per hospital would be required to reach a reliability of 0.8 for distinguishing between hospitals, based on the limited evidence available about the between-hospital variance and other components of variance (see Appendix 2 for the estimates used and methods to project sample size). Holding the total number of reviews constant, increasing the number of reviews per case increases reliability (e.g. 10 reviews per case for 30 cases) more than selecting more cases per hospital (150 cases per hospital with 2 reviews per case). A strategy of only one review per case would provide poor reliability no matter how many total reviews were done per hospital. Figure 2 illustrates how the reliability changes as the numbers of reviews and reviewers per hospital vary.

It is important to emphasize that more extensive and particularly population specific data about the sources of variability in the review procedure could substantially change the projected number of reviews needed in either direction. In general, more heterogeneity across hospitals, more consistent reviewers, evaluating change over time within hospital, or

a focus on relative as opposed to absolute probability of preventable death would result in a more modest and feasible number of reviews needed to produce a reliable estimate.

DISCUSSION

We set out to review the literature on measuring preventable deaths and determine if it would allow us to project how many reviews and reviewers would be required for a/ hospitals to learn lessons from reviewing preventable deaths and b/ for a hospital system to profile hospitals based on their preventable death rates. Secondarily, we looked at whether the literature contained any information on how the reliability of physician implicit review to identify preventable deaths could be improved by refining the measurement procedure. To this end we conducted a review of studies of preventable hospital deaths published from 1980-2015.

The first important finding is that the preventability of death was relatively low in the reviewed studies and this has important consequences for using deaths to assess quality of care. While some studies did vary the probability thresholds and Likert scale anchors for defining preventability as described above, most studies used a similar operational definition of more than a 50/50 chance on balance of probability for defining that a death was preventable. A low prevalence of preventable death should substantially heighten concern about using SMRs calculated from discharge data to profile hospitals. If 95% of deaths are non-preventable, detection of outlier hospitals has an extremely low positive predictive value(3) and any mis-specification of risk adjustment models can also introduce substantial bias.

Another important finding is the lack of any published estimates in the literature of how much variation there is in preventable death rates across hospitals. Without this it is impossible to estimate the reliability for distinguishing between hospitals with respect to their preventable death rates or design an operational program to do so. Using direct measurement, we estimated that as many as 300 or more total reviews could be required per hospital to distinguish between hospitals in a league table with high stakes relegation and promotion consequences. Additionally, holding the total number of reviews per hospitals constant, the optimal number of cases per hospital and reviews per case would require trade-offs to ensure the maximum generalizability and precision.

Furthermore, recall that the purpose of comparing SMRs is explicitly to identify differences

in preventable or avoidable death rates for which the SMR is just a proxy. If the variation in SMRs across hospitals is substantially larger than the variation in preventable death rates as directly measured or the rates are not correlated, it would raise concern that SMRs are measuring something else, most likely unmeasured case-mix differences. Yet, in the absence of this critical piece of information that could support or call into question the validity of SMRs, profiling hospitals based on SMRs is ubiquitous and in the U.S. is tied to significant and increasing financial risk to hospitals.

The literature does provide more data about the reliability of a single measurement to distinguish individual cases with respect to whether a preventable death or preventable adverse event more generally occurred. This reliability estimate is relevant for quality reviews of sentinel cases by hospitals to learn from possible mistakes or for reviews by licensing boards or for cases subject to litigation. It is clear that high reliability is desirable before possible sanctions or changes in procedures are contemplated on the basis of a judgement that a preventable death has occurred. For a typical reliability of 0.35 from the fairly wide range observed, an average of between 8-17 reviewers could be required to reliably distinguish between patients with respect to whether a preventable death occurred. This is far larger than is commonly used for credentialing, legal cases or sentinel case and root cause analysis reviews.

However, providing these specific calculations as examples should not obscure the more important point that different measurement questions and different patient and hospital populations will each require their own estimates of reliability. These reliability estimates can then in turn be used to develop question and population specific calculations of the number of reviewers and reviews per record required so that an estimate with the required precision can be obtained. The numbers may vary substantially with the setting and question.

We also summarise variation in the measurement procedures across studies (Appendix 5). We provide previously unpublished and summary data about many aspects of the procedures used as it was often not reported in the published papers. While the assessment methods had areas in common across the studies, on the whole they were quite

heterogeneous. We found no empirical assessment of how single vs two stage assessments, pro forma tools, reviewer selection or training, reviewer characteristics or environmental influences effect consistency of measurement. Formal reliability or generalizability studies to evaluate different aspects of training and measurement procedures could be built in to an operational program to facilitate improvements in the reliability of measurement. Details of these criteria and methodological issues are discussed as related to existing literature in Appendix 5.

Finally, It is worth re-iterating that the structured implicit case-note review method was originally designed to measure quality, not preventable death, and has a large literature describing its use for this purpose.(58) We should perhaps abandon attempts to measure the absolute proportion of deaths that are preventable as an impossible quest.(13) Physicians are not good at estimating prognostic survival probabilities much less the even more challenging counterfactual probabilities like "what is the probability of survival *if an event had not occurred*" which raises concern about the validity of such estimates.(59-61) Rather, structured implicit review could be used to directly measure the quality of care in the period before a patient's death, in keeping with how these methods were originally designed when developed 30 to 50 years ago.(47, 62-64)

The systematic review component of this study has several limitations. Due to practical reasons, we excluded studies not published in English. We found a large variation in the reported preventable mortality, but with only a limited number of studies we are unable to confirm the exact source of the observed heterogeneity. We have focused on overall hospital mortality and acute general medicines in this review.

CONCLUSIONS

Based on available information, preventable deaths comprise a relatively small fraction of preventable deaths raising concerns about the feasibility of using SMRs as a proxy for preventable deaths. Structured implicit review is a challenging measurement task and it is likely that relatively large numbers of reviews are needed either to allow learning from individual cases or to compare hospitals. However, there is a critical lack of any reported estimates of the hospital variance in preventable death rates which is required to design

systems in a responsible way to profile hospitals based on preventable death rates, whether measured directly or indirectly. There is little evidence about factors that affect the consistency of case-note reviews other than reviewer experience and agreement between reviewers remain poor to moderate.

Any operational system assessing hospital quality around deaths will need to invest in a substantial ongoing effort to quantify the variation across hospitals and reviewers, as well as evaluate how the selection and training of the reviewers and measurement procedures can make the reliability more consistent (see Appendix 5 for an expanded discussion).(65) Attempting to measure preventable deaths on an absolute scale would require engagement with the behavioural science and cognitive psychology literature, pertinent to human and system-wide errors (66) in healthcare,(67) that best locate the 'bounded rationality' of human decision-making,(68) and the biases that plague it.(69, 70) However, whether measuring preventable deaths, or quality more generally, those who want to profile providers must recognize that no program can be designed to distinguish between providers without stable estimates of the amount of variation that exists across those providers.

FIGURE LEGENDS

Figure 1: Review flow diagram of article retrieval and inclusion

Figure 2: Reliability for up to 300 reviews per hospital.

TABLES

Table 1. Characteristics of included studies and methods used for assessing the preventability of deaths or adverse events (AEs)

Author	Location and Date	Target Group / Type of hospital	Grading of preventability*	Threshold for defining a preventable case	% agreement, Kappa (ICC) for preventability	Inter-hospital Variance/ICC	Comments
Dubois, *et al*. 1987.(27, 28)	US. 1985	12 private hospitals	1-4†	≥3† Death as 'probably preventable'	Pearson correlation: 0.71 preventability of **death** (a single independent physician review of 30 charts)	Not reported	• Hospital-wide medical wards with conditions specific to cerebrovascular accident, pneumonia and myocardial infarction. • Acute-care hospitals that were considered outliers with higher and lower than expected mortality • PM estimated from data • 27% of deaths (of all deaths) were preventable
Dubois & Brook. 1988.(26)	US. 1985	12 private hospitals	1-4†	≥3† Death as 'probably preventable'	69%, κ =0.4, 0.3 and 0.2‡ preventability of **death** (182 charts each reviewed by three physicians)	Not Reported	• Hospital-wide medical wards with conditions specific to cerebrovascular accident, pneumonia and myocardial infarction. • Acute-care hospitals that were considered outliers with higher and lower than expected mortality • PM estimated from data • 14% of deaths (of all deaths) were preventable

Brennan, *et al.* 1991.(22)	New York, USA. 1984	51 private and non-federal acute care hospitals	1-6	≥4 negligence is more likely than not	93%, κ =0.24 / preventability of **AE** (based on duplicated review of 318 cases [2/51 hospitals]	Not Reported	• Hospital-wide excluding psychiatric patients. • Non-federal, acute-care hospitals • PM estimated from data. • Weighted figures based on events discovered during index hospitalisation only. • 13.6% of patients with AEs died.
Hayward, *et al*. 1993.(10)	USA, 1988-1990	1 teaching hospital	1-6	≥5 better quality care could have prevented the death	κ =0.5 Death preventable by better quality of care (based on multiple reviews of 34 deaths (80 comparisons) with 122 reviews in total)	N/A (Insufficient Denominator)	• Hospital-wide medical wards with no single diagnostic-related group contributing ≥5% of patient admissions. • Acute-care university teaching hospital • 9% of patient deaths preventable
Best & Cowper. 1994.(21)	USA, 1986	16 Veteran Affair Medical Centers	1-4	≥3 Somewhat likely that better management in the hospital might have prevented patient's death	κ =0.33 'agreement = ≤ 2 positions on 9-point scale' (111 match-pairs from high and low mortality risk Veteran Affairs Medical Centers)	Not Reported	• Veteran Affairs Medical Centers (Small, med/large and Psychiatric/long term types) • 21.6% of patient with better care management might have prevented death (or near the time of death)
Wilson *et al.* 1995.(43)	New South Wales & South Australia. 1992	28 private and public acute care hospitals	1-6	≥4 "Preventability more likely than not, more than 50/50 but close call"	58%, κ =0.33 / preventability of **AE** (based on duplicated review of 6200 cases [all cases positive for screening criteria])	Not Reported	• Hospital-wide excluding day-only admissions & admissions to psychiatric wards. • Preventable AEs and PM estimated from data. • 4.9% of patients with AEs died.

Thomas *et al*. 1999, 2000a, 2000b & 2002.(39-42)	Utah & Colorado, USA. 1992	28 private and public hospitals	1-6	≥4 "More likely than not, >50:50 but close call"	κ =0.19 to 0.23 (95% CI 0.05 to 0.37)) preventability of **AE** (based on 3 independent reviews of 500 records)	Not Reported	• Hospital-wide (13 in Utah and 15 in Colorado) excluding psychiatric and veteran hospitals and patients < 16. • Number of patients with AEs not specified, only total number of AEs. • Based on events discovered during index hospitalisation only • 6.6% of patients with AEs died.
Hayward & Hofer. 2001.(31)	USA. 1994-1995	7 Veterans Affairs hospitals	1-5[§]	≥4[§] "probably" - was the death preventable by optimal care	ICC=0.34 / preventability of **Death** (based on 383 review of 111 cases)	N/A (Insufficient Denominator)	• Hospital-wide excluding data of patients receiving comfort care and non-veterans. • Public Hospitals • Patients with hospital-acquired laboratory abnormality over-sampled • **Reviewed deceased patients only**
Davis *et al.* 2001; 2003.(24, 25) Briant, *et al*. 2006.(23)	New Zealand. 1998	13 public acute care hospitals	1-6	≥4 "Close call, >50:50"	Not reported	Not Reported	• Hospital-wide excluding specialist institutions. • Public hospitals • Over all hospitals there were: 850 AEs; 315 avoidable AEs ≥4; 531 ≥2 • 4.5% of patients with AEs died. • 6.1% of avoidable AEs; unclear concerning disability/death status.

Baker, *et al*. 2004.(20)	Canada. 2000	20 public acute care hospitals	1-6	≥4 "Preventability more than likely (more than 50/50, but close call)"	κ =0.69, (95% CI 0.55-0.83) / preventability of **AE** (based on duplicated review of a random sample of 10% cases)	Not Reported (Hospital Size Groupings preclude *de novo* calculation)	• Hospital-wide excluding psychiatric and obstetric hospitals, day-only admission and patients < 18. • Acute-care hospitals • Weighted percentages to account for total charts per hospital and hospitals per type per province. • 15.7% of patients with AEs died
Michel, *et al.* 2007.(35)	France. 2004	71 private and public hospitals	1-6	≥4 "more likely than not"	67.8%, κ = 0.31 (95% CI 0.05-0.57) / preventability of **AE** (based on 58 cases judged to have AE by both reviewers)	Not Reported	• Hospital wide excluding obstetric hospitals. • Retrospective case-note review & 7-day observation with data collection across 294 wards. • Patients with (preventable) AEs not noted. • 8.2% of patients with AEs died
Soop, *et al*. 2009.(37)	Sweden. 2003-2004	28 public acute care hospitals	1-6	≥4 "more than 50% likelihood"	91%, κ = 0.76 / preventability of **AE** (based on duplicated review of 642 cases [all cases positive for screening criteria])	Not Reported	• Hospital-wide excluding psychiatric, rehabilitation and palliative hospitals and day-only admission. • Acute-care hospitals with high proportion of elderly patients; all deaths occurred in elderly/critically ill patients. • Preventable mortality estimated from data. • 4.1% of patients with AEs died.

Aranaz-Andres, *et al.* 2008(17); 2009(16)	Spain. 2005	24 public hospitals	1-6	≥4 "positive" - not defined	Not reported	Not Reported	• Hospital-wide • Retrospective cohort study. • Patients had 655 AEs; 278 preventable AEs (with at least moderate evidence). • Patients with preventable AEs estimated based on 42.6% of AEs were preventable • Retrospective cohort study. • 4.4% of patients with AEs died. Kappa was reported only for the identification of AEs between reviewers and 'gold standards'.
Aranaz-Andres, *et al.* 2011.(15)	Argentina, Colombia, Costa Rica, Mexico and Peru 2005	58 public hospitals	1-6	≥4 "positive" - not defined	κ ranged from 0.27 to 0.74 between countries / preventability of **AE** (sample size not stated)	Not Reported	• Hospital-wide • Retrospective case-note review and prospective data collection. • Preventable mortality estimated from data • 5.8% of patients with AEs died
Martins, *et al.* 2011.(34)	Brazil, 2003	3 teaching hospitals	1-6	≥4 (wording not described)	Not reported	Not Reported	• Hospital-wide; including obstetric wards. • 38% of patients with AEs died.
Hogan, *et al.* 2012.(32)	England. 2009	10 acute hospitals	1-6	≥4 "Probably preventable, more than 50/50 but close call"	k=0.49 (95% CI 0.2-0.8) / preventability of **Death**, based on duplicated review of 250 cases (25% of sample)	"There were no significant differences between proportions of preventable deaths found at each hospital." (32) p.740	• Hospital-wide excluding obstetric & psychiatric wards, paediatric patients & palliative care. • 100 cases randomly selected from each acute hospital • Reviewed deceased patients only

Sorinola, *et al.* 2012.(38)	England. 2009	1 acute hospital	1-6	≥4 'Preventable death'	κ =0.75 Preventability of **Death** Inter-rater reliability across reviewers of a 'determination of a problem in care' (from sample of 400 notes)	N/A (Insufficient Denominator)	• Hospital-wide excluding obstetric & psychiatric wards, paediatric patients & palliative care. • 400 death cases selected consecutively in 2009 • Preventable mortality estimated from data
Gupta, *et al*. 2013.(30)	US. 2009-2012	1 acute hospital	1-5	≥4 'Possibly preventable'	κ =0.10 Preventability of **Death** agreement between the mortality committee classification and provider classification (Notes sample size not provided)	N/A (Insufficient Denominator)	• Hospital-wide • 2,483 died, 1683 had surveys completed • Preventable mortality estimate provided

Baines, *et al*. 2013; 2015.(18, 19) Zegers, *et al*. 2007; 2009; 2011a; 2011b. (14, 44-46)	The Netherlands 2004 & 2008	33 acute hospitals	1-6	≥4 AE was found to be preventable when the care did not comply with existing professional standards and/or due to shortcomings of a healthcare practitioner, management or system.	**2004** **AE** Positive agreement (+) 54.9% Negative Agreement (-) 66.2% (400 admissions randomly reviewed) **2008** **AE** Positive agreement (+) 54.9% Negative Agreement (-) 82.9% (200 admissions randomly reviewed)	Preventable AEs ICC = 3.7% (hospital-level)	• Hospitals including palliative care and excluding psychiatric, obstetric and paediatric patients • Hospitals were randomly selected on location • Reviewed patients discharged alive and deceased patients • Higher proportion of preventable AEs in deceased than patients discharged alive
Hogan *et al*. 2015. (33) [ll]	England 2012/13	24 acute hospitals	1-6	≥4 Probably avoidable, more than 50-50	κ =0.45 (95% CI 0.24-0.66) / based on random sample of 486 avoidable **Death** cases (grade 4-6)	Not Reported	• Hospitals excluding obstetric, psychiatric and paediatric patients • 100 cases randomly selected from each acute hospital • Reviewed only deceased patients

Manaseki-Holland *et al*. 2016. (13) ¶	England & Wales 2003-2009	22 hospitals	1-5	≥3 On the balance of probability (i.e. > 50% chance)	κ =0.27 (95% CI 0.19-0.39) intra-class correlation across a single review	Not Reported	• Hospitals with inclusion of only respiratory conditions from medical wards • 191 case-notes for those admitted with respiratory complaints and 65 years and over • Case-notes randomly assigned to 2-7 reviewers. (Total of 653 reviews)
Flaatan, *et al*. 2017.(29)	Norway 2011	3 acute hospitals	1-5	≥4 'Possibly preventable'	Not reported	Not reported	• All hospital deaths across 3 hospitals in 2011 (incl. Emergency department) • 1185 deaths notes reviewed across 1 year period • Case notes assigned to six consultant reviewers each from different specialties
Kobewka, *et al*. 2017.(11, 12)	Canada 2013	1 acute hospital	0-100	>50 'Possibly preventable'	Reliability=0.68** ICC=0.14 (480 deaths each reviewed by 4 reviewers)	N/A (Insufficient Denominator)	• Hospital excluding paediatrics • 480 deceased case-notes (structure case abstracts) produced across 3-month admission period • Case-notes randomly assigned to 4 physician reviewers
Roberts, *et al*. 2017.(36)	UK 2012-2015	4 North-East England, UK acute care Trusts	1-6 (PRISM) 1-5 (NCEPOD)	≥4 (PRISM) ≥3 (NCEPOD)	κ=0.45 (PRISM) κ=N/A (NCEPOD) (Not provided)	N/A	• All hospital deaths across 4 Trusts • 7370 medical records reviewed • Case notes reviewed predominantly by consultants, some by nurses.

* Scale of degree of preventability. This tends to range from "6, *(virtually) certain evidence of preventability"* through to "1 *(virtually) no evidence for preventability".*

† We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1 *(definitely) preventable death* through to 4 *(definitely not) preventable deaths.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

‡ For cerebrovascular accident, myocardial infarction and pneumonia, respectively.

§ We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1 *(definitely) preventable death* through to 5 *(definitely not) preventable deaths.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

‖ "In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in healthcare had not occurred?"

¶ The 'England' study has been extracted from the 2016 paper as the 'US' data has been included in Hofer & Hayward (2001). We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1 *(definitely) preventable death* through to 5 *(definitely not) preventable deaths.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

** Mean reliability across four reviewers.

Table 2. Summary of study processes and review methods

Category		No.	References
Inclusion of a Screening Stage	No screening stage	4	(32, 33, 36, 37)
	Yes (16-18), criteria)	15	(10, 14-26, 31, 34, 35, 38-46)
	Trigger tool	4	(15, 26, 34, 38)
Scale used for implicit judgement	Binary	0	
	4 point Likert	2	(21, 26)
	5 point Likert	3	(13, 31, 36)
	6 point Likert	16	(10, 14-20, 22-25, 32-46)
	Continuous	2	(11-13)
Reviewer Screening Stage 1	Physician	7	(13, 14, 18, 19, 27-29, 32, 33, 36, 44-46)
	Nurse	11	(14-19, 21-25, 34, 35, 37-42, 44-46)
	Pharmacist	1	(38)
Reviewer Review Stage 2	Physician Expert Advice Available	15	(14-25, 27, 28, 34-46)
	Pharmacist support	0	
	Nurse support	0	
Duration of Expert Advice	Indefinite duration	3	(10, 33, 36)
	Temporary duration	3	(16, 17, 21, 23-25)
	No stated duration	2	(13, 33)
Reviewer affiliations	External to the institution being reviewed	20	(10-26, 31-35, 37-46)*
	Internal	2	(21, 36)*
Hospital Anonymisation	Undertaken	5	(13, 23-25, 31-33)
	NOT undertaken	17	(10-12, 14-22, 26-28, 34-46)
Clinical Experience of Physicians	< 5 years	0	
	5-10 years	4	(11, 12, 15-17, 20)
	> 10 years	7	(21, 32-34, 36, 37, 43)
	Previous experience not mentioned	2	(10, 39-42)
	No mention of experience	5	(22-28, 35)
Speciality of Physicians	General medicine/Internal medicine (Alone)	13	(10, 15-17, 20-25, 32, 34, 35, 37, 38, 43)
	Internal medicine & specialists	9	(11-14, 18, 19, 21, 26, 31, 33, 36, 39-42, 44-46)
Review Discrepancies and Disagreements Reconciled	Physicians	3	(14, 18, 19, 36, 43-46)
	Nurses	0	
	Medical Health Analysts / Records Analyst	1	(22)
	Executive Board	2	(16, 17, 37)
	Information not available	6	(20, 21, 23-28, 39-42)
Physician Reviewer Training Duration	≤ 1 day	7	(14, 18, 19, 21, 23-25, 27, 28, 32, 33, 38, 44-46)
	1 - 3 days	7	(13, 20, 31, 34, 36, 39-43)
	≥ 3 days	3	(16, 17, 35, 37)
	Not stated	4	(10-12, 15, 26)
Training Content	Case-note Exposure	12	(10, 13, 14, 18-28, 31, 36, 37, 44-46)
	Specialist Advice Provided	8	(14, 16-19, 21, 23-25, 27, 28, 31, 32, 36, 44-46)
	Absence of Preventability definition	18	(10, 13-20, 22-26, 31-35, 37-46)
	Familiarity with study tools	14	(10, 13, 14, 18-25, 27, 28, 33, 34, 36-42, 44-46)

** Best, et al. (1994) (21) was half external and half internal.*

Table 3. Preventable mortality and/or adverse events reported in the included studies

Author	No. of admitted patient case notes reviewed (with or without initial screening) [a]	No. of deceased patient case notes reviewed [b]	No. of admission case notes selected after screening for review [c]	Preventable AEs (% of admissions)	Preventable AEs (% of all AEs [c])	Preventable mortality (% of admissions [a])	Preventable mortality (% of deceased [b])	Threshold for preventability & comments
Dubois, et al. 1987. (28, 53)(USA)	NR	182	NR	NR	NR	NR	26.9% 49/182	Preventability score ≥3 out of 4*
Dubois & Brook. 1988. (26)(USA)	1,946	182	NR	NR	NR	4.6% 90/1,946	26.9% 49/182 14% 25/182	Preventability score ≥3 out of 4* (majority decision) Preventability score ≥3 out of 4* (unanimous decision)
Brennan, et al. 1991. (22)(USA)	30,121	NR	7,743	306 (1.02% weighted)	3.96% 306/7743	0.30% 89/30, 121	NR	Causation score ≥1 on a 0-6 scale Preventability score ≥ 4 out of 6
Hayward, et al. 1993. (10)(USA)	675	NR	NR	NR	NR	0.44% 3/675	9% 3/34	Preventability score ≥4 out of 6
Best & Cowper. 1994. (21)(USA)	NR	222***	NR	NR	NR	NR	21.6% median	Preventability score ≥3 out of 4"
Wilson, et al. 1995. (43)(Australia)	14,179	114	1,718	1,205 (8.50%)†	NR	0.55% 78/14,179	29.00%	Causation score ≥2 out of 6 Preventability score ≥4 out of 6

Thomas, et al. 1999, 2000a, 2000b & 2002. (39-42)(USA)	14,700	NR	448	3.00% 448/14, 700	NR	0.265% 39/14,700	NR	Causation score ≥4 out of 6; Preventability: 'an adverse event was considered preventable if it was avoidable by any means currently available unless that means was not considered standard care.' The implicit judgement methods are similar to those used in Bates, et al. 1997.[70]
Hayward & Hofer. 2001.(31) (USA)	NA	111	NA	NA	NR	0.23% - 0.61% (at least possibly preventable) (95% CI)	22.7%; 6.0%	Preventability score ≥3 out of 5§ Preventability score ≥4 out of 5§ (preventable scores weighted to account for over-sampling of patients with laboratory abnormality)
Davis, et al. 2001; 2003. Briant, et al. 2006. (23, 24)(New Zealand)	6,579	118	850	6.28% 413/6,579	48.6% 413/850	0.36% 24/6579	19.8 – 20.7%	Preventability score Not Reported (NR) Causation score ≥2 out of 6 Preventability score ≥2 out of 6
Baker, et al. 2004. (54)‡ (Canada)	3,692	236	1512	2.8% (95% CI 2.0% to 3.6%)^	7.01% 106/1,512	0.66% (95% CI 0.37% -0.95%)^ (death from preventable AE)	16.9% 40/236§	Causation score ^{3}4 out of 6 Preventability score ≥4 out of 6 (Preventability score Not Reported (NR) §)
Michel, et al. 2007. (35)(France)	8,754	NR	NR	1.08% 95/8,754	NR	0.09% 8/8,754	NA	Preventability score Not Reported (NR) Causation score ≥4 out of 6 Preventability score ≥4 out of 6
Soop, et al. 2009. (37)(Sweden)	1,967	10	241	8.6% 169/1,967	70.1% 169/241	0.25% 5/1,967	NR	Causation score ≥4 out of 6
Aranaz-Andres, et al. 2008; 2009. (16, 17)(Spain)	5,624	225	1,755	11.65% 655/5,624	37.3% 655/1,755	0.07% 5/5,624¶	4.5%	Causation score ≥4 out of 6 Preventability score ≥4 out of 6

Aranaz-Andres, et al. 2011.(15) (Argentina, Colombia, Costa Rica, Mexico and Peru)	11,379	NR	1,754	10.47% 1,191/11,379	67.9% 11,91/1,754	0.32 780*0.058 72?/11,379	NR	Causation score ≥4 out of 6 Preventability score ≥4 out of 6
Martins, et al. 2011. (34)(Brazil)	1,103	94	1,103	5.07% 56/1,103	5.07% 56/1,103	8.50%	26.6% 25/94	Causation score ≥4 out of 6 Preventability score ≥4 out of 6
Hogan, et al. 2012.** (32)(England)	NR	1000	NR	NR	NR	NR	5.2% 52/1,000	Preventability score ≥4 out of 6 (reporting 1 of 3)
Sorinola, et al. 2012. (38)(England)	NR	400	NR	NR	NR	NR	3.5% 14/400	Preventability score ≥4 out of 6
Gupta, et al. 2013. (30)(USA)	NR	1,683	NR	NR	NR	NR	2.50% 42/1,683	Preventability score ≥4 out of 5
Baines, et al. 2013; 2015. Zegers, et al. (18, 19, 44-46)2007; 2009; 2011a; 2011b. (The Netherlands)	11,949	762	1,130	NR	NR	NR	4.5%	Preventability score ≥4 out of 6
Hogan, et al. 2015.(33) †† (England)	NR	2,400	NR	NR	NR	NR	3% 101/2,400	Preventability score ≥4 out of 6
Manaseki-Holland, et al. 2016. (13)(England)	NR	191[µ]	NR	NR	NR	NR	10% (median) Q1 3% Q3 28%	Preventability score ≤2 out of 5

Flaatten, et al. 2017. (29)(Norway)	59,605	1,167	NR	NR	NR	0.057% 34/59,605	2.91% 34/1,167	Preventability score ≥ 50 out of 100
Kobekwa, et al. 2017. (11, 12)(Canada)	14,267	480[µ]	NR	NR	NR	0.22% 31/14,267	6.46% 31/480[‡‡]	Preventability score ≥ 50 out of 100††
Roberts, et al. 2017. (36)(UK)	NR	7,194	NR	NR	NR	NR	0.47% 34/7,194	Preventability score ≥ 50 out of 100

Causation score is the score given to the likelihood of the adverse event being caused by medical care/management: causation score ≥2 out of 6 corresponds to 'at least slight to modest evidence of management causation'; causation score ≥4 out of 6 corresponds to 'management causation more likely – more than 50/50'.

Preventability score ≥2 out of 6 corresponds to 'at least slight to modest evidence of preventability'. Preventability score ≥4 out of 6 corresponds to 'preventability more than likely – more than 50/50'.

[*] We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1 *(definitely) preventable death* through to 4 *(definitely not) preventable deaths.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[†] This indicator is for deaths considered with a high level of preventability.

[‡] Of all Adverse events = 459 50% probability of membership in the "possibly preventable" class.

[§] Figures are taken from direct author response rather than published data.

[^] Adjusted for sampling frame.

[¶] Associated with preventable AE.

[**] "Was the patient's death due to problems in the healthcare or did problems in healthcare contribute to the death?"

*** Pairs were matched across high observed-to-expected mortality ~(OTEM) and low OTEM Veteran affairs hospitals

[††] "In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in healthcare had not occurred?"

[‡‡] >50% probability of membership in the "possibly preventable" class

[µ] multiple reviews were undertaken with the case notes

NA = Not Assessed; NR = Not Reported

REFERENCES

1. Intelligence, Dr Foster. The Dr Foster Hospital Guide 2009: How Safe is Your Hospital. London: Dr Foster Intelligence. 2009.
2. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. Bmj. 2010;340:c2016.
3. Girling AJ, Hofer TP, Wu J, Chilton PJ, Nicholl JP, Mohammed MA, et al. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. BMJ quality & safety. 2012;21(12):1052-6.
4. Scott IA, Brand CA, Phelps GE, Barker AL, Cameron PA. Using hospital standardised mortality ratios to assess quality of care—proceed with extreme caution. Med J Aust. 2011;194(12):645-8.
5. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. Journal of epidemiology and community health. 2007;61(11):1010-3.
6. Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, et al. Inter-rater reliability of case-note audit: A systematic review. Journal of Health Services Research and Policy. 2007;12(3):173-80.
7. Tender for Retrospective case record review programme - Lot 2 Tender Documentation for a National Programme to use retrospective record review to estimate the proportion of hospital deaths due to problems in care in English acute NHS trusts [press release]. 2015.
8. National Quality Board, UK. National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting, Investigating and Learning from Deaths in Care. 2017.
9. The Royal College of Physicians. RCP hosts pioneering national Mortality Case Record Review Programme 2016 [cited 2016 29/04/2016]. Available from: https://www.rcplondon.ac.uk/news/rcp-hosts-pioneering-national-mortality-case-record-review-programme.
10. Hayward RA, McMahon Jr LF, Bernard AM. Evaluating the care of general medicine inpatients: How good is implicit review? Annals of Internal Medicine. 1993;118(7):550-6.
11. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. The prevalence of potentially preventable deaths in an acute care hospital: A retrospective cohort. Medicine (United States). 2017;96(8).
12. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. APPENDICES - The prevalence of potentially preventable deaths in an acute care hospital: A retrospective cohort: . 2017.
13. Manaseki-Holland S, Lilford RJ, Bishop JR, Girling AJ, Chen Y-F, Chilton PJ, et al. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. BMJ Quality & Safety. 2016.
14. Zegers M, de Bruijne MC, Spreeuwenberg P, Wagner C, van der Wal G, Groenewegen PP. Variation in the rates of adverse events between hospitals and hospital departments. International Journal for Quality in Health Care. 2011;23(2):126-33.
15. Aranaz-Andrés JM, Aibar-Remón C, Limón-Ramírez R, Amarilla A, Restrepo FR, Urroz O, et al. Prevalence of adverse events in the hospitals of five Latin American countries: Results of the 'Iberoamerican study of adverse events' (IBEAS). BMJ Quality and Safety. 2011;20(12):1043-51.
16. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Burillo J, Requena-Puche J, Terol-García E, Kelley E, et al. Impact and preventability of adverse events in Spanish public hospitals: Results of the Spanish national study of adverse events (ENEAS). International Journal for Quality in Health Care. 2009;21(6).
17. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Murillo J, Ruiz-López P, Limón-Ramírez R, Terol-García E. Incidence of adverse events related to health care in Spain: Results of the Spanish National Study of Adverse Events. Journal of Epidemiology and Community Health. 2008;62(12):1022-9.
18. Baines RJ, Langelaan M, De Bruijne MC, Asscheman H, Spreeuwenberg P, Van De Steeg L, et al. Changes in adverse event rates in hospitals over time: A longitudinal retrospective patient record review study. BMJ Quality and Safety. 2013;22(4):290-8.
19. Baines RJ, Langelaan M, De Bruijne MC, Wagner C. Is researching adverse events in hospital deaths a good way to describe patient safety in hospitals: A retrospective patient record review study. BMJ Open. 2015;5(7).

20. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. Canadian medical association journal. 2004;170(11):1678-86.
21. Best WR, Cowper DC. The Ratio of Observed-to-Expected Mortality as a Quality of Care Indicator in Non-Surgical VA Patients. Medical Care. 1994;32(4):390-400.
22. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard medical practice study I. New England Journal of Medicine. 1991;324(6):370-6.
23. Briant R, Buchanan J, Lay-Yee R, Davis P. Representative case series from New Zealand public hospital admissions in 1998 - III: Adverse events and death. New Zealand Medical Journal. 2006;119(1231).
24. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals II: Preventability and clinical context. New Zealand Medical Journal. 2003;116(1183).
25. Davis P, Lay-Yee R, Briant R, Schug S, Scott A, Johnson S, et al. Adverse events in New Zealand public hospitals: principal findings from a national survey. 2001.
26. Dubois RW, Brook RH. Preventable deaths: Who, how often, and why? Annals of Internal Medicine. 1988;109(7):582-9.
27. Dubois RW, Rogers WH, Draper D, Brook RH. Author's Reply: Does Hospital Mortality Predict Quality? New England Journal of Medicine. 1988;318(24):1622.
28. Dubois RW, Rogers WH, Moxley III JH, Draper D, Brook RH. Hospital inpatient mortality. New England Journal of Medicine. 1987;317(26):1674-80.
29. Flaatten H, Brattebø G, Alme B, Berge K, Rosland JH, Viste A, et al. Adverse events and in-hospital mortality: an analysis of all deaths in a Norwegian health trust during 2011. BMC health services research. 2017;17(1):465.
30. Gupta M, Fuchs B, Cutilli C, Cintolo J, Reinke C, Kean C, et al. Preventable mortality: does the perspective matter when determining preventability? Journal of Surgical Research. 2013;184(1):54-60.
31. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: Preventability is in the eye of the reviewer. Journal of the American Medical Association. 2001;286(4):415-20.
32. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: A retrospective case record review study. BMJ Quality and Safety. 2012;21(9):737-45.
33. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: Retrospective case record review and regression analysis. BMJ (Online). 2015;351.
34. Martins M, Travassos C, Mendes W, Pavão ALB. Hospital deaths and adverse events in Brazil. BMC Health Services Research. 2011;11.
35. Michel P, Quenon JL, Djihoud A, Tricaud-Vialle S, De Sarasqueta AM. French national survey of inpatient adverse events prospectively assessed with ward staff. Quality and Safety in Health Care. 2007;16(5):369-77.
36. Roberts AP, Morrow G, Walkley M, Flavell L, Phillips T, Sykes E, et al. From research to practice: results of 7300 mortality retrospective case record reviews in four acute hospitals in the North-East of England. BMJ Open Quality. 2017;6(2).
37. Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: A retrospective medical record review study. International Journal for Quality in Health Care. 2009;21(4):285-91.
38. Sorinola OO, Weerasinghe C, Brown R. Preventable hospital mortality: learning from retrospective case record review. JRSM short reports. 2012;3(11):77.
39. Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: Population based review of medical records. British Medical Journal. 2000;320(7237):741-4.

40. Thomas EJ, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Annals of Internal Medicine. 2002;136(11):812-6.
41. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Medical Care. 2000;38(3):261-71.
42. Thomas EJ, Studdert DM, Newhouse JP, Zbar BIW, Howard KM, Williams EJ, et al. Costs of medical injuries in Utah and Colorado. Inquiry. 1999;36(3):255-64.
43. Wilson McL R, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. Medical Journal of Australia. 1995;163(9):458-71.
44. Zegers M, Bruijne MC, Wagner C, Groenewegen PP, Waaijman R, Wal G. Design of a retrospective patient record study on the occurrence of adverse events among patients in Dutch hospitals. BMC Health Services Research. 2007;7(1):1.
45. Zegers M, de Bruijne MC, de Keizer B, Merten H, Groenewegen PP, van der Wal G, et al. The incidence, root-causes, and outcomes of adverse events in surgical units: Implication for potential prevention strategies. Patient Safety in Surgery. 2011;5(1).
46. Zegers M, De Bruijne MC, Wagner C, Hoonhout LHF, Waaijman R, Smits M, et al. Adverse events and potentially preventable deaths in Dutch hospitals: Results of a retrospective patient record review study. Quality and Safety in Health Care. 2009;18(4):297-302.
47. Rubenstein LV, Kahn KL, Harrison E, Sherwood MJ, Rogers WH, Brook RH. Structured Implicit Review of the Medical Record. 1991.
48. Pearson ML, Lee JL, Chang BL, Elliott M, Kahn KL, Rubenstein LV. Structured implicit review: a new method for monitoring nursing care quality. Medical care. 2000:1074-91.
49. Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk-adjusted surgical outcomes: chart review of process of care. International Journal for Quality in Health Care. 2001;13(3):187-96.
50. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? Jama. 1992;267(17):2349-54.
51. Fischhoff B. Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human perception and performance. 1975;1(3):288.
52. Fischhoff B. Hindsight≠ foresight: the effect of outcome knowledge on judgment under uncertainty. Quality and Safety in Health Care. 2003;12(4):304-11.
53. Brett AS, Roberts M, Ruderman AJ, Levin MH, Pine M, Smith D, et al. Does Hospital Mortality Predict Quality? New England Journal of Medicine. 1988;318(24):1622-4.
54. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: The incidence of adverse events among hospital patients in Canada. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2004;170(11):1678-86.
55. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychosocial Measurement, 20, 37-46. 1960.
56. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. American journal of epidemiology. 1987;126(2):161-9.
57. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. Statistics in medicine. 2002;21(14):2109-29.
58. Kahn KL, Rubenstein LV, Sherwood MJ, Brook RH. Structured Implicit Review for Physician Implicit Measurement of Quality of Care. 1989.
59. Chow E, Harth T, Hruby G, Finkelstein J, Wu J, Danjoux C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally iII cancer patients? A systematic review. Clinical Oncology. 2001;13(3):209-18.
60. Christakis NA, Smith JL, Parkes CM, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort studyCommentary: Why do doctors overestimate? Commentary: Prognoses should be based on proved indices not intuition. BMJ. 2000;320(7233):469-73.

61. Glare P, Virik K, Jones M, Hudson M, Eychmuller S, Simes J, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. BMJ. 2003;327(7408):195.
62. Empire State Medical, Scientific and Educational Foundation, Inc. Rochester region perinatal study. Medical review project, Empire State Medical, Scientific and Educational Foundation, Inc. New York state journal of medicine. 1967;67(9):1205-10.
63. Richardson FM. Peer review of medical care. Med Care. 1972;10(1):29-39.
64. Sheps MC. Approaches to the quality of hospital care. Public health reports (Washington, DC : 1896). 1955;70(9):877-86.
65. Hogan H. The problem with preventable deaths. BMJ quality & safety. 2015:bmjqs-2015-004983.
66. Reason J. Human error: Cambridge University Press; 1990.
67. Reason J. Beyond the organisational accident: the need for “error wisdom” on the frontline. Quality and Safety in Health Care. 2004;13(suppl 2):ii28-ii33.
68. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. American psychologist. 2003;58(9):697.
69. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Medical Informatics and Decision Making. 2016;16(1):138.
70. Blumenthal-Barby J, Krieger H. Cognitive Biases and Heuristics in Medical Decision Making A Critical Review Using a Systematic Search Strategy. Medical Decision Making. 2014:0272989X14547740.